CR

Prediction of Ligand–Receptor Binding Thermodynamics by Free Energy Force Field Three-Dimensional Quantitative Structure–Activity Relationship Analysis: Applications to a Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase

Prabha Venkatarangan† and Anton J. Hopfinger*

Laboratory of Molecular Modeling and Design (M/C-781), College of Pharmacy, University of Illinois at Chicago, 833 South Wood Street, Chicago, Illinois 60612-7231

Received September 8, 1998

Glucose analogue inhibitors of glycogen phosphorylase, GP, may be of clinical interest in the regulation of glycogen metabolism in diabetes. The receptor geometry of glycogen phosphorylase *b*, GP*b*, is available for structure-based design and also for the evaluation of the thermodynamics of ligand–receptor binding. Free energy force field (FEFF) 3D-QSAR analysis was used to construct ligand–receptor binding models. FEFF terms involved in binding are represented by a modified first-generation AMBER force field combined with a hydration shell solvation model. The FEFF terms are then treated as independent variables in the development of 3D-QSAR models by correlating these energy terms with experimental binding energies for a training set of inhibitors. The genetic function approximation, employing both multiple linear regression and partial least squares regression data fitting, was used to develop the FEFF 3D-QSAR models for the binding process and to scale the free energy force field for this particular ligand–receptor system. The significant FEFF energy terms in the resulting 3D-QSAR models include the intramolecular vacuum energy of the unbound ligand, the intermolecular ligand–receptor van der Waals interaction energy, and the van der Waals energy of the bound ligand. Other terms, such as the change in the stretching energy of the receptor on binding, change in the solvation energy of the system on binding, and the change in the solvation energy of the ligand on binding are also found in the set of significant FEFF 3D-QSAR models. Overall, the binding of this class of ligands to GP*b* is largely characterized by how well the ligand can sterically fit into the active site of the enzyme. The FEFF 3D-QSAR models can be used to estimate the binding free energy of any new analogue in substituted glucose series prior to synthesis and testing.

Introduction

Structure-based design currently involves the use of the structure of the receptor and/or, if available, the structure of a ligand–receptor complex to perform molecular modeling studies to elucidate the features of ligand–receptor binding. The three-dimensional structure of receptors and, increasingly, of ligand–receptor complexes are available through advances in molecular biology, protein expression and purification, X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy. Unfortunately, the accurate and reliable prediction of the thermodynamics of ligand–receptor binding has remained problematic. That is, the computational equivalent of the in vitro binding assay has remained an elusive goal.

The correlation between binding affinities and a set of two-(2D) and three-dimensional (3D) descriptors for a series of ligands has been used to develop quantitative structure–activity relationships (QSARs).[1] For those cases in which the receptor geometry is known, the physicochemical properties derived from ligand–receptor interactions can potentially be used in building receptor-dependent (RD) 3D-QSARs. Hopfinger and coworkers in 1981 reported the correlation of biological activities with log *P*, the water–octanol partition coefficient of the ligand, and the *calculated* intermolecular binding affinities for anticancer anthracyclines intercalating between adjacent DNA base pairs as a 3D-QSAR.[2] Holloway et al. have constructed a regression equation which relates calculated interaction energies for HIV-1 protease inhibitor complexes and the corresponding observed in vitro enzyme inhibition.[3] Ortiz et al. reported a method termed comparative binding energy (COMBINE) analysis for constructing RD 3D-QSARs in which the receptor geometry is used in computing the candidate QSAR descriptors.[4] A recent paper by Tokarski and Hopfinger reported a RD 3D-QSAR methodology termed free energy force field (FEFF) 3D-QSAR analysis as applied to a set of peptidomimetic renin inhibitors.[5] Succinctly, the principal features of FEFF 3D-QSAR analysis are (1) all of the enthalpy and entropy contributions to the ligand–receptor interaction in a solvent medium are taken into consideration, (2) the set of enthalpy and entropy contributions to binding are treated as the independent variables in developing a 3D-QSAR model for ligand–receptor binding, and (3) the optimal FEFF 3D-QSAR

† Current Address: Pharmacopeia Inc., 101 College Road East, Princeton, NJ 08540.

10.1021/jm980515p CCC: $18.00
Published on Web 06/02/1999

Table 1. Breakdown of the FEFF Interaction Terms, XY, for a Ligand (L) and a Receptor (R) in a Medium (M)

binding component(s)	type of interaction energy, XY	change in internal energy, symbols	change in entropy, symbols
ligand L	intramolecular ligand conformational energy LL	$\Delta E_L(LL) = E_{LR}(LL) - E_L(LL)$	$\Delta S_L(LL) = S_{LR}(LL) - S_L(LL)$
ligand L	ligand solvation energy LM	$\Delta E_L(LM) = E_{LR}(LM) - E_L(LM)$	$\Delta S_L(LM) = S_{LR}(LM) - S_L(LM)$
solvent medium M	solvent reorganizational energy MM	$\Delta E_M(MM) = E_{LR}(MM) - [E_L(MM) + E_R(MM)]$	$\Delta S_M(MM) = S_{LR}(MM) - [S_L(MM) + S_R(MM)]$
receptor R	intramolecular receptor conformational energy RR	$\Delta E_R(RR) = E_{LR}(RR) - E_R(RR)$	$\Delta S_R(RR) = S_{LR}(RR) - S_R(RR)$
receptor R	receptor solvation energy RM	$\Delta E_R(RM) = E_{LR}(RM) - E_R(RM)$	$\Delta S_R(RM) = S_{LR}(RM) - S_R(RM)$
ligand–receptor RL	intermolecular ligand–receptor energy LR	$\Delta E_{LR}(LR) = E_{LR}(LR)$	$\Delta S_{LR}(LR) = S_{LR}(LR)$

model is constructed using a genetic algorithm. FEFF 3D-QSAR analysis is the intermolecular extension of molecular shape analysis (MSA)[6] and can be viewed as a simulation of an in vitro binding assay.

Ligand–receptor molecular dynamics simulations (MDSs) are used in the FEFF method to assemble the ligand–receptor binding states. The smaller the chemical system (less atoms), the more computationally economical is the corresponding MDS. One of the goals of the work reported in this paper is to devise a procedure to reduce the size of the receptor model to facilitate shorter MDSs while retaining reliable results.

This paper reports the FEFF 3D-QSAR analysis of a set of glucose analogue inhibitors of glycogen phosphorylase *b* (GP*b*). Glycogen is the carbohydrate reserve of most metabolically active cells in mammals. GP catalyzes the first step in the phosphorolysis of glycogen to glucose-1-phosphate. In muscle, glycolysis of glucose-1-phosphate provides energy to sustain muscle contraction. The liver converts glucose to provide fuel for other tissues. GP exists in two interconvertible states through reversible phosphorylation, the inactive *b* form (predominantly T state) and the active *a* form (predominantly R state). Hepatic glycogen metabolism is regulated by glucose through promotion of inactivation of GP*a*.[7] Glucose inactivates GP*a* by competitive inhibition of glucose-1-phosphate and stabilizes the inactive T state.

Methods

1. The FEFF 3D-QSAR Formalism. The ligand–receptor interaction can be expressed as

$$\underset{\text{(unbound state)}}{\text{L·M} + \text{R·M}} \overset{K}{\rightleftharpoons} \underset{\text{(bound state)}}{\text{(LR)·M}} \quad (1)$$

where L is the ligand, R is the receptor, M is the solvent medium, and K is the inhibition constant expressed relative to LR dissociation. The difference in free energy between the bound and unbound states of a ligand, L, to a receptor, R, in a solvent medium, M, can be stated as

$$\Delta G^\circ = G^\circ_{LR} - (G^\circ_L + G^\circ_R) = -RT \ln K \quad (2)$$

where ΔG° is the *standard* binding free energy, G°_{LR} is the free energy of the bound or complex state, G°_L is the free energy of the unbound ligand, G°_R is the free energy of the unbound receptor, R is the gas constant, and T is the temperature of the system. The free energy of an enzyme–ligand complex can be approximately broken down into a set of component interactions as follows

$$G^\circ_{LR} = [G_{LR}(LL) + G_{LR}(RR) + G_{LR}(MM) + G_{LR}(LR) + G_{LR}(LM) + G_{LR}(RM)] \quad (3)$$

where $G_{LR}(XY)$ refers to the interaction between X and Y.

The interaction terms can be divided into their respective enthalpy, H°_{LR}, and entropy, S°_{LR}, contributions.

$$G^\circ_{LR} = H^\circ_{LR} - TS^\circ_{LR} \quad (4)$$

At low solute concentration the enthalpy terms, $H_{LR}(XY)$, can be represented by their respective internal energies, $E_{LR}(XY)$,

$$H^\circ_{LR} = E^\circ_{LR} = [E_{LR}(LL) + E_{LR}(RR) + E_{LR}(MM) + E_{LR}(LR) + E_{LR}(LM) + E_{LR}(RM)] \quad (5)$$

and the entropy term, $S_{LR}(XY)$, contributions as

$$S_{LR} = [S_{LR}(LL) + S_{LR}(RR) + S_{LR}(MM) + S_{LR}(LR) + S_{LR}(LM) + S_{LR}(RM)] \quad (6)$$

The unbound ligand, G°_L, and receptor, G°_R, free energies have the following components

$$G^\circ_L = [G_L(LL) + G_L(LM) + G_L(MM)] \quad (7)$$

$$G^\circ_R = [G_R(RR) + G_R(RM) + G_R(MM)] \quad (8)$$

The enthalpy contributions of L and R at low concentration, $H_L(XY)$ and $H_R(XY)$, can also be represented by their internal energies, $E_L(XY)$ and $E_R(XY)$, as in eq 5. The complete set of contributions to the internal and entropy energies and their representations is given in Table 1.

The terms in Table 1 can be the independent variables used in the FEFF 3D-QSAR analysis. However, the free energy of binding, ΔG, can also be represented by the individual free energy force field terms for L, R, and LR in Table 1 along with their respective weighting (regression) coefficients, α_i. This representation for ΔG, eq 9, can be used to provide additional descriptors for FEFF 3D-QSAR analysis

$$\Delta G = \alpha_1 \Delta E_{stretch} + \alpha_2 \Delta E_{bend} + \alpha_3 \Delta E_{torsion} + \alpha_4 \Delta E_{vdW} + \alpha_5 \Delta E_{electrostatic} + \alpha_6 \Delta E_{hydrogen-bonding} + \alpha_7 \Delta E_{solvation} + \alpha_8 T\Delta S \quad (9)$$

where $\Delta E_{stretch}$ is the unbound to bound change in

internal energy for bond stretching, ΔE_{bend} is the change in bond angle bending energy, $\Delta E_{torsion}$ is the change in torsional energy, ΔE_{vdW} is the change in van der Waals interaction energy, $\Delta E_{electrostatic}$ is the change in electrostatics interaction energy, $\Delta E_{hydrogen-bonding}$ is the change in hydrogen-bonding energy, $\Delta E_{solvation}$ is the change in solvation energy, and ΔS is the change in the entropy of the L, R, M system which can be partitioned into component contributions. The hydration shell model proposed by Hopfinger[8] was included in the potential energy function to calculate the solvation energies.

Koehler and Hopfinger have applied a group additive property (GAP) method to calculate the conformational entropies of linear chain polymers.[9] The assumption of the GAP concept is that some intrinsic contribution to any composite property, $P(i)$, is associated with the ith structural group of the molecule. The composite property, P, is simply taken to be the sum of the $P(i)$ composing the molecule. The GAP method of Hopfinger and Koehler is called torsion angle unit, TAU, theory. A TAU is defined by adjacent structural units connected by a bond about which some torsion angle θ occurs.

The application of the TAU theory to ligand–receptor systems was reported by Tokarski and Hopfinger.[5] This method assumes torsional conformational entropy makes the largest entropic contribution to the ligand–receptor system and estimates the component values. The TAU values[9] are selected for the torsion angle types found in the receptor and inhibitors, and the sum of the appropriate TAU entropy values allows an estimation of the corresponding ligand and receptor conformational entropies.

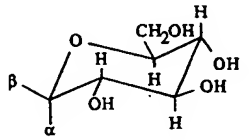
In this study, the glucose analogue inhibitors have quite limited conformational flexibility. Still, the TAU method was applied to a subset of the ligands in the training set and ligand conformational entropy estimated. Variation in conformational entropy was found, as expected, to be small, and the TAU entropies did not appear in the optimized FEFF 3D-QSAR models for this subset. It was, therefore, assumed that the ligand contribution from a change in torsional conformational entropy upon binding to the receptor is small and nearly constant across the analogue series. Further, the change in the torsional conformational entropy of the receptor upon ligand binding was assumed to be constant across the analogue series. No attempt was made to estimate entropy from the ensemble of states sampled in the MDS by constructing the corresponding partition function.

The internal energy change upon ligand–receptor binding is given by

$$\Delta E_X = E_{LR,X} - (E_{L,X} + E_{R,X}) \qquad (10)$$

where X represents each of the internal energy contributions as defined in eq 9. The potential function parameters used to calculate the nonbonded, electrostatic, torsional, bond stretching, and bond angle bending energy terms of eq 9 were taken from the AMBER force field.[10] Missing force field parameters (torsional, bond stretching, and bond angle bending) were scaled from a set proposed by Hopfinger[11] and the MM2 force field.[12] The set of atoms most similar to those of the missing AMBER parameter is identified for a parameter which has both AMBER and MM2 (or Hopfinger) values.

Table 2. Structure–Activity Data for the Glucose Analogue Inhibitors of Glycogen Phosphorylase *b* Used in the FEFF 3D-QSAR Training Set



compd	α	β	K_i (mM)	ΔG_{303} (kcal/mol)
1	H	NHC(=O)CH_3	0.032	6.23
2	H	NHC(=O)CH_2CH_3	0.039	6.11
3	H	NHC(=O)CH_2Br	0.044	6.04
4	H	NHC(=O)CH_2Cl	0.045	6.03
5	H	NHC(=O)C_6H_5	0.081	5.67
6	H	NHC(=O)$CH_2CH_2CH_3$	0.094	5.58
7	H	NHC(=O)NH_2	0.14	5.34
8	H	C(=O)$NHCH_3$	0.16	5.26
9	H	NHC(=O)CH_2NH_2	0.37	4.76
10	C(=O)NH_2	H	0.37	4.76
11	H	C(=O)NH_2	0.44	4.65
12	H	C(=O)$NHNH_2$	0.40	4.17
13	H	SH	1.00	4.16
14	CH_2OH	H	1.50	3.92
15	OH	H	1.70	3.84
16	H	C(=O)NHC_6H_5	5.40	3.14
17	H	OH	7.40	2.95
18	H	CH_2CN	9.00	2.84
19	OH	CH_2OH	15.80	2.50
20	H	OCH_3	24.70	2.23
21	CH_2NH_2	H	34.50	2.03
22	C(=O)$NHCH_3$	H	36.70	1.99
23	CH_3	H	53.10	1.77

The ratio of the known parameter from the AMBER and MM2 force fields is determined. The unknown AMBER parameter value is then scaled by the same ratio against the known MM2 value. This linear scaling approximation in force field parametrization is further compensated by the subsequent force field fitting process which is central to the FEFF methodology and has been described above.

2. Inhibitory Binding Constants–Dependent Variables. A training set of glucose analogue inhibitors of GP*b* were reported along with their inhibitory binding constants (K_i).[13-16] The kinetic binding studies were performed as described by Martin and co-workers.[13] The structure–activity data for the training set of glucose analogue inhibitors of GP*b* are given in Table 2. The K_i values are expressed as millimolar (mM) units. The ΔG values are calculated from the K_i values using eq 2.

3. Receptor Geometry. Martin and co-workers cocrystallized glucose bound to GP*b* at 2.4 Å resolution.[13] The coordinates of this complex were obtained from the Brookhaven Protein Data Bank[18] under the PDB entry 2GP*b*. The starting protein structure used in the FEFF 3D-QSAR analysis was the refined structure of the glucose complex.[13] Hydrogens were added to the PDB structure through an option in the QUANTA modeling package.[19] Water molecules located in the crystal structure of glucose–GP*b* complex were not included in the FEFF protein model. Steric contact violations were identified by 3000 steps of MDS using the MOLSIM package.[20] The bad steric interactions were relieved by perturbation conformational scans over the side chains of the residues causing them. AMBER partial atomic charges[10] were assigned to all atoms of

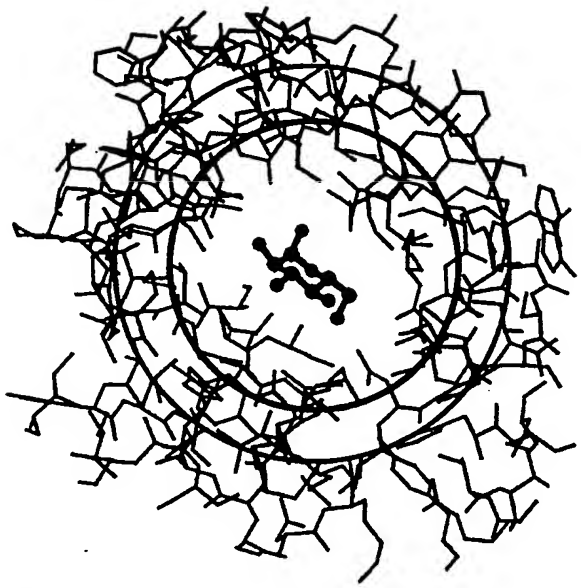


Figure 1. Schematic representation of the geometry used to determine the size of the model enzyme. The amino acid residues centered around the ligand (thick line) are included in the 8 Å model size (inner concentric circle). The outer circle represents the 10 Å model.

the enzyme structure. The potentially ionizable residues of the enzyme were modeled as neutral.

4. Building and Docking the Ligands. The geometry of bound glucose was used as the starting structure for building the ligand analogues of Table 2. The Chemlab-II molecular modeling program[21] was used to add substituents to the glucose ring. Substituent geometries were optimized by fixed valence geometry conformational analysis. CNDO/2 charges were assigned to the ligands. The conformation and alignment of the reference crystal glucose ring structure was used to dock the glucose analogues. Bad steric contacts between a few of the ligand analogues and side chains of enzyme residues were relieved during initial docking.

5. Enzyme Model Size Determination. The complete enzyme model contains 13470 atoms which includes protons. This large enzyme size would require unrealistically extensive MDSs to yield meaningful results. Thus, the GP*b* enzyme system was scaled down to economize the MDSs. The largest inhibitor (compound **16** of Table 2) was docked at the active site to determine the minimum size of an effective enzyme model needed to do reliable FEFF 3D-QSAR analysis. The derivation of the enzyme binding model was done using the "pruning" method of Tokarski and Hopfinger.[5] Spherical enzyme models of 12, 10, and 8 Å radii, centered around the docked ligand, see Figure 1, were examined in order to prune the enzyme without losing binding information. The pruned enzyme models essentially consist of amino acid residues clustered around the active site. Localized conformational changes of the residues at the active site were seen for the binding of some analogues.[13-16] The pruned models are designed to allow for these conformational changes at and near the active site.

The residues that had at least one non-hydrogen atom within the pruning sphere were included in the corresponding enzyme model. The pruning process usually leads to an enzyme model consisting of a number of nonbonded (unconnected) peptide residue clusters. Peptide residue clusters separated by less than five intervening amino acid residues were "connected" by including the intervening amino acid residues. This approach is intended to retain local geometric integrity of the enzyme model in the pruning process.

The enzyme model size was evaluated for both conformational and binding integrity by performing MDSs on the scaled down trial models. A MDS of 2 ps at 300 K was performed on each of the model enzymes of different pair-interaction size using a nonbonded cutoff of 16 Å. The molecular dielectric was 3.5. The modified AMBER force field was used in the MDSs. Comparisons of the total intermolecular van der Waals and electrostatic energies, respectively, for each of the three different size enzyme models were made.

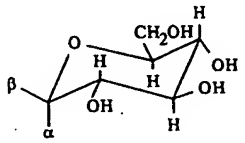
A root-mean-square (RMS) fit of the lowest energy structure of each of the three model enzymes to the starting crystal structure was made. The RMS fit of the whole enzyme structure from the MDS with respect to the crystal structure was also made. The ligand–receptor interaction energies were calculated for each of the enzyme models.

Some atoms of each enzyme model had to be constrained to prevent major deviations from the starting crystal structure. The enzyme models contain unconnected peptides, but assigning fictitious high masses to all of the atoms in an enzyme model provides "momentum reservoirs" which can impart equivalent structural and dynamic effects as the "absent" amino acid residues of the complete enzyme. Different enzyme models were explored by assigning a fictitious mass of 1000, 2500, and 5000, respectively, to each atom of an enzyme model and then evaluating the RMS fit to the initial X-ray crystal structure as described by Tokarski and Hopfinger.[5] Inhibitors 1, 16, and 23 of Table 2 were used in this enzyme pruning and evaluation study. A MDS of 10 ps, with a step size of 0.5 fs, was performed for each of the trial enzyme mass models with a bound ligand at 300 K. The maintenance of the structural integrity of the complex, as measured by the RMS fit, was used to evaluate and select the preferred fictitious mass value. It should be pointed out that the conformational entropy terms are estimated using the TAU method which is independent of the masses assigned in this enzyme pruning scheme.

6. Sampling Temperature Determination. The temperature of an MDS can be only approximate to the corresponding real temperature of the chemical system. The specific representation of the force field used in a MDS can also influence the relationship between actual and simulation temperatures. Thus, temperature was considered as a scaling variable in this FEFF 3D-QSAR analysis.

The temperature sampling schedule scheme developed by Tokarski and Hopfinger[5] for the bound and unbound states of the ligand enzyme system, using a subset of the analogue training set, see Table 3, was carried out at 400, 350, 300, 200, 100, 50, 25, 10, and 5 K. "Mini" FEFF 3D-QSAR models derived from the ligand subset were used to determine the preferred simulation temperature. That is, the preferred MDS

Table 3. Subset of the Glucose Analogue Inhibitor Training Set Used in the Determination of the Optimum MDS Sampling Temperature



compd	α	β	K_i (mM)	ΔG_{303} (kcal/mol)
1	H	NHC(=O)CH_3	0.032	6.23
4	H	NHC(=O)CH_2Cl	0.045	6.03
5	H	NHC(=O)C_6H_5	0.081	5.67
8	H	C(=O)$NHCH_3$	0.16	5.26
15	OH	H	1.70	3.84
16	H	C(=O)NHC_6H_5	5.40	3.14
21	CH_2NH_2	H	34.50	2.03
22	C(=O)$NHCH_3$	H	36.70	1.99
23	CH_3	H	53.10	1.77

Table 4. r^2 and $xv - r^2$ Values of the Best FEFF 3D-QSARs Realized from the MDSs at Different Simulation Temperatures

temp (K)	r^2	$xv - r^2$
400	0.56	0.38
350	0.62	0.49
300	0.85	0.79
200	0.70	0.56
100	0.45	0.30
50	0.54	0.33
25	0.55	0.45
10	0.33	0.23
5	0.50	0.29

temperature corresponds to the best fit of FEFF thermodynamic parameters to the experimental binding free energies. The correlation coefficient, r^2, and leave-one-out cross validation coefficient, $xv-r^2$, of the best FEFF 3D-QSAR for each simulation temperature are given in Table 4.

7. Computational Details. The binding simulation sampling scheme was initiated by an MDS of 20 ps, using a time step of 0.5 fs, on the ligand–enzyme complex model. The structures of the models and their corresponding FEFF energy terms were recorded every 0.2 ps of the simulation. Modeling of the unbound state consisted of isolating both the bound ligand and the corresponding receptor from the lowest energy geometry of their complex realized from the bound state MDS modeling. A MDS of the unbound ligand was then performed at 300 K for 100 ps using a time step of 1 fs. A corresponding MDS of the unbound receptor was performed for a sampling time of 20 ps, with a step size of 0.5 fs, at 300 K. The hydration shell model[8] was used to calculate the solvation energy of the lowest energy conformation obtained from the MDS of the complex, the unbound ligand, and the unbound receptor. The lowest energy geometry of the complex, E_{LR}, unbound ligand, E_L, and unbound receptor, E_R, were used, respectively, to obtain the corresponding free energy force field terms as described in Table 1 and eq 9. The ensemble averaged free energy force field terms were also calculated using the trajectories of MDSs as the ensembles.

8. Construction of the FEFF 3D-QSAR Models. The nonscaled FEFF energy terms were used as descriptors (independent variables), and the genetic function approximation (GFA) optimization method,[22] employing multiple linear regression, MLR, was used to construct trial QSAR models. The robustness of each model was tested by evaluating statistical measures of fit which included r^2, $xv-r^2$, the F-statistic, F, and the lack-of-fit, LOF.[22,23] The number of possible FEFF descriptors is large compared to the number of analogues in the training set. The LOF measure in the GFA prevents overfitting of data by assigning a penalty for the addition of independent variables to a model.

To determine if the top five 3D-QSAR models provide common, or distinct, structure–activity information, the correlation coefficients of the residuals in the error (observed activity − predicted activity) between pairs of models were computed. Equivalent models are expected to have identical distributions in the residuals of error. Distinct models should have noncorrelated patterns in their residuals of fit (error). This type of analysis has been suggested by Rogers[24,25] as a diagnostic to determine the subset of distinct models among a set of good models realized in a GFA analysis.

Results

The size of the enzyme model was selected based on a combination of geometric stability and the variance of the ligand–enzyme interaction energy for the 12, 10, and 8 Å model radii. The ligand–model enzyme interaction energies showed no significant differences among the three model sizes evaluated for an atom-pair nonbonded cutoff of 16 Å. The atom RMS fit of each lowest energy geometry from the corresponding MDS for the 12, 10, and 8 Å radii models to the enzyme crystal structure was 1.06, 1.04, and 1.29 Å, respectively. The total ligand–enzyme interaction energies for the 12, 10, and 8 Å models are 44.57, 45.01, and 44.26 kcal/mol, respectively. The RMS fit of the whole enzyme structure from the MDS with respect to the crystal structure was 1.30 Å. The enzyme model size of 10 Å was selected for all FEFF final MDSs as a compromise between structural integrity and computational efficiency.

A fictitious mass assignment to all non-hydrogen atoms of the enzyme model was found to be a minimum necessary condition to maintain the geometric integrity of the enzyme model. Moreover, a fictitious mass of 2500 assigned to each of the atoms in the 10 Å enzyme model provided optimal geometric model stability (Figure 2). In addition, studies on the role of MDS schedule indicated that longer simulations did not necessarily lead to lower energy structures. However, MDSs of more than 20 ps often caused the enzyme model to diverge substantially from the crystal geometry. It was found that a simulation temperature of 300 K provided a trial set of FEFF terms which when used as descriptors yielded the best FEFF 3D-QSAR model. It should be noted that the bioassays for measuring binding were performed at 303 K.

Overall, a sampling temperature of 300 K on the 10 Å enzyme model, with a fictitious mass of 2500 assigned to all enzyme model atoms, and a MDS of 20 ps provided the best compromise between computational time and the realization of low-energy states of geometries close to the crystal structure. The low-energy structures, obtained from the MDSs, of all the ligand–enzyme systems of the training set were found to have RMS fits

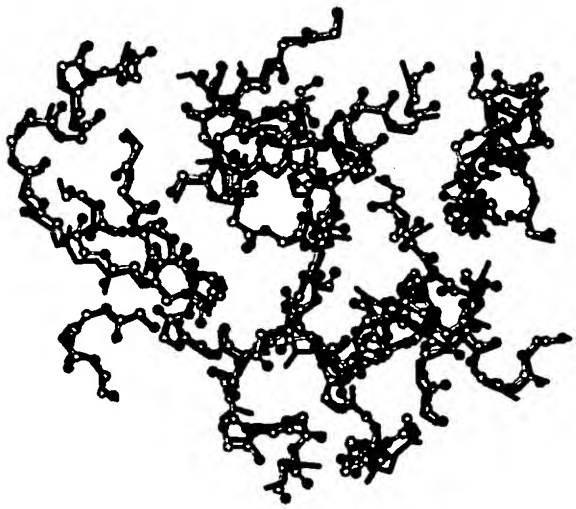


Figure 2. Overlap of the protein backbone of the initial crystal structure (dark lines) with the low-energy model enzyme structure after molecular dynamics simulation at 300 K using a heavy mass assignment of 2500 to each atom of the 10 Å model.

to the crystal structure of less than 1.5 Å (Figures 3 and 4).

The trial FEFF 3D-QSAR models were constructed using the GFA-MLR option of the WOLF program[26] for a sample size of 23 analogues. The models derived from GFA-MLR analysis are listed in Table 5 along with their statistical measures of fit and robustness as discussed earlier. A smoothing factor of 0.5 was found to optimize model size (number of independent variables) and predictiveness.

The top five FEFF 3D-QSAR models based on the applied statistical measures of fit are reported in Table 5. The FEFF terms found as significant descriptors in the GFA-MLR analysis are defined in Table 6. All ligands in the training set (no outliers) are used in constructing the models in Table 5. The van der Waals intermolecular ligand−receptor binding energy, $E_{LR,vdW}$, the intramolecular energy of the unbound ligand, E_L-(LL), and the van der Waals energy of bound ligand, $E_{LR,vdW}$(LL), are found in all five top models and are not correlated to other significant FEFF terms found among the top five FEFF 3D-QSAR models. The stretching energy of the bound ligand, $E_{LR,stretch}$(LL), the electrostatic energy of the bound ligand, $E_{LR,electrostatic}$(LL), and the hydrogen-bonding energy of the unbound ligand, $E_{L,hb}$(LL), are present interchangeably in the five top models. The variables $E_{LR,el}$, ΔE_{hb}, $\Delta E_{electrostatic}$, see Table 6, although not found in the FEFF 3D-QSAR models,

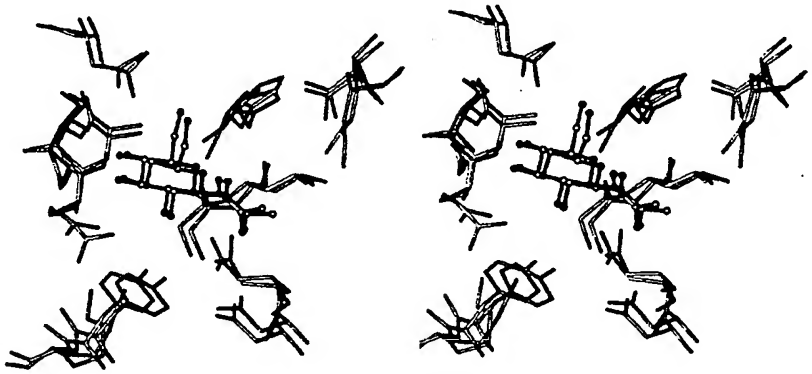


Figure 3. A stereoview of the overlap of the active site amino acid residues of the enzyme crystal structure with the low-energy representative structure of the model enzyme after molecular dynamics simulation at 300 K with a heavy mass of 2500 assigned to each atom of the 10 Å model with the α-methyl analogue (compound **23**) bound.

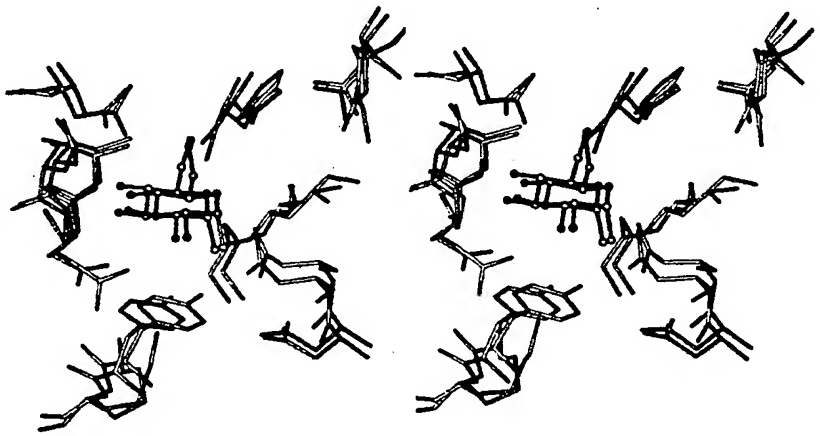


Figure 4. A stereoview of the overlap of the active site amino acid residues of the enzyme crystal structure with the low-energy representative structure of the model enzyme after molecular dynamics simulation at 300 K with a heavy mass of 2500 assigned to each atom of the 10 Å model with the β-methylacetyl analogue (compound **1**) bound.

Table 5. Top Five FEFF 3D-QSAR Models for the Prediction of ΔG for the Complete Training Set Listed in Table 2

3D-QSAR	r^2 [a]	$xv - r^2$ [b]	F [c]	LOF [d]
$\Delta G = -0.10E_{L}(LL) - 0.16E_{LR,vdW} + 0.07E_{LR,stretch}(LL) - 0.80E_{LR,vdW}(LL) - 0.31$	0.72	0.58	11.3	1.10
$\Delta G = -0.10E_{L}(LL) - 0.17E_{LR,vdW} + 0.06\Delta E_{L,stretch}(LL) - 0.79E_{LR,vdW}(LL) - 0.02$	0.71	0.57	10.9	1.13
$\Delta G = -0.13E_{L}(LL) - 0.26E_{LR,vdW} + 0.08E_{LR,stretch}(LL) - 0.74\Delta E_{L,vdW}(LL) - 0.83$	0.71	0.53	10.9	1.13
$\Delta G = -0.10E_{L}(LL) - 0.19E_{LR,vdW} + 0.08E_{LR,electrostatic}(LL) - 0.74E_{LR,vdW}(LL) - 0.61$	0.70	0.55	10.7	1.14
$\Delta G = -0.10E_{L}(LL) - 0.19E_{LR,vdW} + 0.02E_{L,hb}(LL) - 0.74E_{LR,vdW}(LL) - 0.68$	0.70	0.53	10.6	1.16

[a] Correlation coefficient. [b] Cross-validated r^2 (leave-one-out method).[29] [c] F-statistic. [d] LOF is the lack-of-fit measure.[22,23]

Table 6. Definitions of the Significant FEFF Terms Found in the Top FEFF 3D-QSAR Models

$E_{L}(LL)$	the intramolecular energy of the unbound ligand
$E_{LR,vdW}$	the van der Waals intermolecular ligand–receptor energy
$\Delta E_{L,vdW}(LL)$	the change in the van der Waals energy of the ligand upon binding
$E_{LR,vdW}(LL)$	the intramolecular van der Waals energy of the bound ligand
$E_{L,vdW}(LL)$	the intramolecular van der Waals energy of the unbound ligand
$E_{L,stretch}(LL)$	the stretching energy of the unbound ligand
$E_{LR,stretch}(LL)$	the stretching energy of the bound ligand
$\Delta E_{L,stretch}(LL)$	the change in the stretching energy of the ligand on binding
$\Delta E_{R,stretch}(RR)$	the change in the stretching energy of the receptor on binding
$\Delta E_{R,bend}(RR)$	the change in the bending energy of the receptor on binding
$\Delta E_{stretch}$	the change in the stretching energy on binding
$E_{LR,electro}(LL)$	the intramolecular electrostatic energy of the bound ligand
$E_{L,hb}(LL)$	the intramolecular hydrogen-bonding energy of the unbound ligand
$E_{LR,el}$	the electrostatic intermolecular ligand–receptor energy
ΔE_{hb}	change in hydrogen-bonding energy of the whole system upon binding
$\Delta E_{electrostatic}$	change in electrostatic energy of the whole system upon binding
$E_{L}(LM)$	the solvation energy of the unbound ligand
$\Delta E_{L}(LM)$	change in solvation energy of the ligand upon binding
ΔE_{solv}	change in solvation energy of the whole system upon binding

are highly correlated with some of the significant FEFF terms in the top models, see Table 7. The descriptors within any given model are not significantly correlated to one another.

Partial least squares (PLS) regression yielded a three-component FEFF model as being most significant. All the top MLR models given in Table 5 contain the three descriptors identified by PLS, i.e., the van der Waals intermolecular ligand–receptor binding energy, $E_{LR,vdW}$, the intramolecular energy of the unbound ligand, E_{L}-(LL), and the van der Waals energy of bound ligand, $E_{LR,vdW}(LL)$. Thus, the FEFF terms used in the top models are likely the major thermodynamic properties governing the relative binding thermodynamics. The relative contributions of the individual FEFF descriptors during the GFA model optimization are shown by the crossover versus descriptor usage plot in Figure 5. Once again the LOF measure prevents overfitting of the data and is particularly appropriate in the FEFF 3D-QSAR approach since many of the energy terms are derived from the same source and the number of energy terms (the trial set of independent variables) is large compared to the number of observations (analogues in the training set). A statistically poor 3D-QSAR model ($r^2 = 0.3$, $xv - r^2 = 0.12$) was obtained when the biological activity column was randomized, suggesting that the FEFF 3D-QSAR models in Table 5 were not due to random correlations.

To determine if the top five 3D-QSAR models are providing common, or distinct, structure–activity information, the correlation coefficients of the residuals in error (observed activity − predicted activity) between pairs of models were computed and are reported in Table 8. All of the top five models are highly correlated to one another, indicating there is only one unique FEFF 3D-QSAR model which is selected as the model with the highest $xv - r^2$ value, namely model 1 in Table 5.

Outliers were defined as those analogues whose difference in observed and predicted ΔG values are greater than 2 standard deviations from the mean. The resulting two outliers for the training set are compounds **18** and **19** (Table 2) as can be seen in Figure 6. The removal of these two outliers and GFA-MLR refitting on the remaining data set of 21 compounds yielded the FEFF 3D-QSAR models shown in Table 9. In addition to the dispersion terms found in the original best model of Table 5, the total LR(M) change in the solvation energy upon binding, ΔE_{solv}, the change in the solvation energy of the ligand on binding, $\Delta E_{L}(LM)$, and the unbound ligand solvation energy, $E_{L}(LM)$, are now found as significant descriptors in the best "outlier-free" four-descriptor FEFF 3D-QSAR models. The stretching, electrostatic, and hydrogen-bonding energy terms of the models given in Table 5 are replaced by the change in the stretching energy upon binding, $\Delta E_{stretch}$, and the change in the stretching energy of the receptor upon ligand binding, $\Delta E_{stretch}(RR)$.

GFA-MLR optimization of the FEFF 3D-QSAR models was determined by developing three-, four-, and five-descriptor families of models (Table 9). This type of multiple model representation helps to identify the origin and significance of each term in the manifold of FEFF 3D-QSAR models. It is possible that the five-term models may constitute an overfit case, but these models should be considered within the context of the manifold set of models in Table 9. The three-descriptor models are predominantly comprised of the dispersion energy terms and the intramolecular vacuum energy of the unbound ligand. The solvation energy term is added to the four-descriptor models, and this increases the r^2 value from 0.77 to 0.88. The five-descriptor models do not provide a better "explanation" (significantly increased data fits) as compared to the four-descriptor models and are found to be comprised of the same terms found in the three- and four-descriptor models.

Compound **18**, the β-cyanomethyl analogue, contains a rigid side chain, and compound **19** is the only compound in the analogue series, other than glucose, to have an α-hydroxy substituent. These singular properties may be the sources of why these two analogues are outliers. The N atom of the cyano group of compound **18** and the O of the α-hydroxy group of compound **19**

Table 7. Linear Cross-Correlation Matrix of the Descriptors in the Top FEFF 3D-QSAR Models[a]

energy terms[b]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.00																
2	−0.30	1.00															
3	0.08	−0.31	1.00														
4	0.02	−0.22	0.97	1.00													
5	0.14	0.08	−0.35	−0.45	1.00												
6	0.16	−0.02	0.10	0.01	0.89	1.00											
7	0.06	0.03	0.18	0.19	−0.11	−0.03	1.00										
8	0.13	0.72	−0.31	−0.26	0.09	−0.03	0.13	1.00									
9	−0.04	−0.50	0.38	0.34	−0.16	0.00	0.68	−0.64	1.00								
10	0.26	0.10	0.13	0.17	0.02	0.11	0.05	0.09	−0.03	1.00							
11	−0.26	−0.10	−0.14	−0.17	−0.02	−0.11	−0.05	−0.09	0.03	−1.00	1.00						
12	−0.02	−0.26	0.21	0.13	0.20	0.29	0.13	−0.15	0.21	−0.77	0.77	1.00					
13	−0.06	−0.06	0.08	0.05	0.16	0.21	−0.05	−0.20	0.11	0.37	−0.37	−0.24	1.00				
14	0.02	0.05	−0.14	−0.09	−0.15	−0.21	−0.02	0.19	−0.16	−0.35	0.35	0.21	−0.99	1.00			
15	−0.03	−0.04	−0.06	−0.04	−0.14	−0.18	−0.01	0.07	−0.06	−0.47	0.47	0.34	−0.98	0.98	1.00		
16	0.04	−0.03	0.14	0.11	0.00	0.06	−0.09	−0.17	0.06	0.45	−0.45	−0.40	0.92	−0.92	−0.94	1.00	
17	−0.05	0.03	−0.16	−0.12	−0.01	−0.07	0.10	0.18	−0.05	−0.45	0.45	0.41	−0.92	0.92	0.93	−1.00	1.00

[a] The matrix includes descriptors that are correlated to those FEFF descriptors determined from the top five models using the GFA-MLR scheme. [b] 1 = E_L(LL), 2 = $E_{LR,vdW}$, 3 = $E_{LR,stretch}$(LL), 4 = $\Delta E_{L,stretch}$(LL), 5 = $\Delta E_{R,stretch}$(RR), 6 = $\Delta E_{stretch}$, 7 = $E_{LR,vdW}$(LL), 8 = $E_{L,vdW}$(LL), 9 = $\Delta E_{L,vdW}$(LL), 10 = E_L(LM), 11 = ΔE_L(LM), 12 = ΔE_{solv}, 13 = $E_{L,hb}$(LL), 14 = $E_{LR,electrostatic}$(LL), 15 = $E_{LR,el}$, 16 = ΔE_{hb}, 17 = $\Delta E_{electrostatic}$.

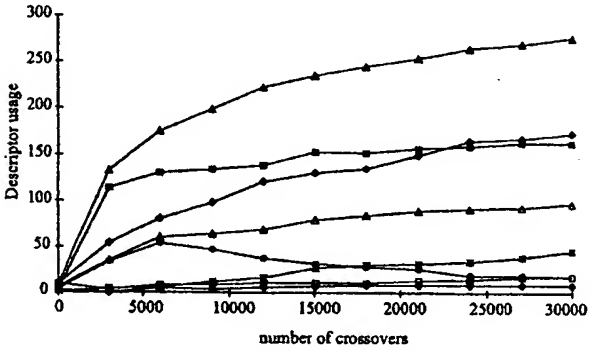


Figure 5. A plot of the GFA-MLR optimization for predicting ΔG. Descriptor usage plotted as a function of number of crossovers in the GFA-MLR analysis is shown here. Symbols: ◇, E_L(LL) use; ■, $E_{LR,vdW}$(LL) use; ▲, $E_{LR,vdW}$ use; △, $E_{L,vdW}$-(LL) use; *, $E_{LR,stretch}$(LL) use; ○, $E_{L,hb}$(LL) use; □, $\Delta E_{L,stretch}$-(LL) use; ◆ $E_{LR,electrostatic}$(LL) use. The FEFF terms (descriptors) are defined in Table 6.

Table 8. Linear Correlation Matrix of the Residuals of Error for the Top Five FEFF 3D-QSAR Models from GFA-MLR Optimization

	model 1	model 2	model 3	model 4	model 5
model 1	1.00				
model 2	1.00	1.00			
model 3	0.89	0.89	1.00		
model 4	0.91	0.89	0.85	1.00	
model 5	0.91	0.90	0.87	0.92	1.00

were also found to form hydrogen bonds to the active site residues through a water molecule (water bridges) (ref). The only source of hydrogen bonding for these substituent groups to the active site residues seem to be through such water bridges.[14] Water-bridge ligand–receptor hydrogen bonding also occurs for compounds **10** and **15** which are found to have relatively large (within the outlier definition) residual values. Ligand–receptor hydrogen bonding via water molecule bridges is not explicitly considered in our FEFF MDSs. Therefore, poor modeling of water-bridge ligand–receptor interactions may be another source for generating outliers.

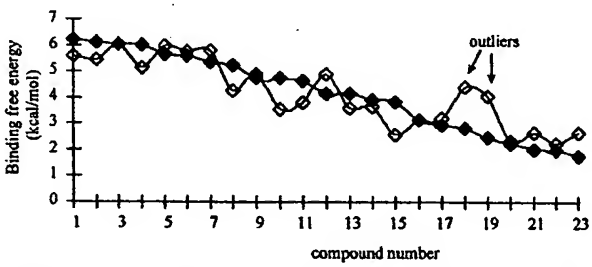


Figure 6. Plot of the observed (◆) and predicted (◇) (using model 1 of Table 5) binding free energy, ΔG, versus the compound number (Table 2). The outliers are identified in the plot.

Discussion

The ligand–receptor modeling approximations employed in this FEFF 3D-QSAR study are listed in Table 10. This table also describes how each approximation influences the calculations, and suggestions are given as to how to minimize the adverse impacts of the approximations on the development of FEFF 3D-QSAR models.

Explicit water molecules have not been considered in the MDSs. Rather, a hydration shell model to estimate solvation energetics has been included in the force field. The hydration shell model, when applied over the course of a MDS, can produce major distortions in molecular geometry. This behavior arises because the force derived from the hydration shell potential function is unrealistically large and operates over the entire length of the MDS time step. To minimize this flaw in the representation of the force field solvation term, the FEFF solvation energies are only computed for selective, and representative, low-energy geometries along a MDS trajectory.

The absence of explicit water molecules during the MDS eliminates the possibility of forming specific types of solute–solvent structures, such as a water bridge between the bound ligand and the enzyme as is observed in the crystal complexes for some analogues in this data set. On the other hand, assignment of explicit waters in an MDS requires very long MDSs in order to

Table 9. Top Five Three-, Four-, and Five-Descriptor FEFF 3D-QSAR Models for Prediction of ΔG after Deletion of Outliers

3D-QSARs	r^2 [a]	$xv-r^2$ [b]	F [c]	LOF [d]
Three-Descriptor Models				
$\Delta G = -0.10E_L(LL) - 0.16E_{LR,vdW} - 0.85E_{LR,vdW}(LL) + 0.45$	0.77	0.66	20.45	1.20
$\Delta G = -0.70E_{L,vdW}(LL) - 0.06\Delta E_{R,stretch}(RR) - 1.02E_{LR,vdW}(LL) + 0.82$	0.75	0.63	20.21	1.30
$\Delta G = -0.85E_{L,vdW}(LL) - 0.04\Delta E_{R,bend}(RR) - 1.15E_{LR,vdW}(LL) + 0.76$	0.73	0.60	20.20	1.35
$\Delta G = -0.10E_{LR,vdW} - 0.06\Delta E_{R,stretch}(RR) - 1.06E_{LR,vdW}(LL) - 0.52$	0.72	0.60	19.63	1.40
$\Delta G = -0.14E_L(LL) - 0.26E_{LR,vdW} - 0.69\Delta E_{L,vdW}(LL) + 0.01$	0.71	0.51	20.89	1.40
Four-Descriptor Models				
$\Delta G = -0.09E_L(LL) - 0.14E_{LR,vdW} - 0.05\Delta E_{R,stretch}(RR) - 0.99E_{LR,vdW}(LL) + 0.08$	0.88	0.80	11.12	0.90
$\Delta G = -0.10E_L(LL) - 0.15E_{LR,vdW} - 0.04\Delta E_{stretch} - 0.91E_{LR,vdW}(LL) + 0.04$	0.83	0.73	12.22	1.10
$\Delta G = -0.42E_L(LM) - 0.65E_{L,vdW}(LL) - 0.06\Delta E_{R,stretch}(RR) - 1.01E_{LR,vdW}(LL) - 1.51$	0.83	0.73	10.21	1.21
$\Delta G = 0.40\Delta E_L(LM) - 0.65E_{L,vdW}(LL) - 0.06\Delta E_{R,stretch}(RR) - 1.01E_{LR,vdW}(LL) - 1.41$	0.82	0.72	10.10	1.29
$\Delta G = 0.29\Delta E_{solv} - 0.61E_{L,vdW}(LL) - 0.07\Delta E_{R,stretch}(RR) - 1.10E_{LR,vdW}(LL) - 1.36$	0.82	0.72	9.95	1.31
Five-Descriptor Models				
$\Delta G = -0.08E_L(LL) - 0.12E_{LR,vdW} - 0.06\Delta E_{R,stretch}(RR) - 1.05E_{LR,vdW}(LL) + 0.19\Delta E_{solv} - 1.11$	0.90	0.80	13.56	0.76
$\Delta G = -0.09E_L(LL) - 0.12E_{LR,vdW} - 0.06\Delta E_{R,stretch}(RR) - 0.99E_{LR,vdW}(LL) + 0.19E_{L,stretch}(LL) - 1.05$	0.90	0.81	14.56	0.77
$\Delta G = -0.08E_L(LL) - 0.13E_{LR,vdW} - 0.05\Delta E_{R,stretch}(RR) - 0.99E_{LR,vdW}(LL) + 0.22E_L(LM) - 1.13$	0.89	0.79	13.25	0.79
$\Delta G = -0.09E_L(LL) - 0.13E_{LR,vdW} - 0.04\Delta E_{R,stretch}(RR) - 1.03E_{LR,vdW}(LL) + 0.19E_{LR,stretch}(LL) - 0.001$	0.89	0.81	13.10	0.80
$\Delta G = -0.08E_L(LL) - 0.13E_{LR,vdW} - 0.08\Delta E_{R,stretch}(RR) - 1.01E_{LR,vdW}(LL) + 0.03\Delta E_{stretch} + 0.21$	0.88	0.80	12.22	0.85

[a] Correlation coefficient. [b] Cross-validated r^2 (leave-one-out method (ref)). [c] F-statistic. [d] LOF is the lack-of-fit measure (ref).

Table 10. Molecular Modeling Approximations, Their Impact on Modeling, and the Approaches Used in This Study To Minimize Approximation Impact

approximations	impact on modeling	approaches to minimize impact
1. FEFF representation		
a. solvation energies estimated using a hydration shell model	incorrect balance between solvation energy and the rest of the FEFF during MDSs	consider only the solvation energies for low-energy states, or the conformer state used to construct the QSAR
b. explicit water molecules are not included	hydrogen bonding of ligand through water molecules (water bridges) to active site residues	examine outliers within context of missing explicit water interactions
c. entropic contributions are considered to be constant for the inhibitor analogue series	neglect of conformational flexibility on binding	entropy contributions can be estimated, if necessary, by a group additive model and scaled with respect to temperature in the FEFF 3D-QSAR fitting procedure
2. the LR, L, and R are modeled as being neutral	multiple protonation states are possible and could influence electrostatic energetics	the protonation state held constant for the entire training set, so error should be "constant" over the training set; a neutral state approximates solvation and counterion effects on FEFF interactions
3. scaled down receptor model	the scaled down receptor geometry can deviate from the crystal geometries over a long MDS and some RR and LR interactions are eliminated	heavy masses assigned to each of the atoms of the scaled down model to model missing momentum reservoir of the rest of the enzyme
4. MDS temperature	balance the enthalpy and entropic contributions to ΔG	the preferred MDS temperature corresponds to the best FEFF 3D-QSAR model of a subset of the training set
5. sampling schemes used to explore the geometry-energy states of the LR, R, and L	the sampling schemes may be incomplete with respect to sampling bound and unbound ligand conformations and to monitoring the change in geometry of the receptor for the bound and unbound states	use experimental data for bound ligand alignment and ligand–receptor geometry for defining the bound and unbound ligand reference states

adequately sample the states of the system and build up a meaningful equilibrium profile. In essence, such explicit water MDSs on inhibitor–enzyme systems are not practical, and the incomplete ensemble sampling associated with such attempted simulations very likely yields large errors in the estimation of the solvation energetics.

The inhibitors are relatively small, rigid, and high analogues to one another so that changes in intramolecular ligand conformational entropy in the binding processes are small and can be neglected in the FEFF. Likewise, changes in the receptor geometry appear to be small from the X-ray structures of free and bound enzymes. Thus, entropy changes due to the receptor geometry have not been considered in the development of the FEFF 3D-QSAR models. Only the solvation entropy inherent to the hydration shell solvation model is resident in the development of the QSAR models.

Van der Waal and stretching energy terms dominate as the key descriptors (binding terms) in both the whole training set and the "outlier-free" best FEFF 3D-QSAR models. This finding suggests that the steric fitting of the inhibitor into the receptor active site is the critical feature that distinguishes binding among the inhibitors in the training set. Moreover, the steric fitting and corresponding van der Waals and valence geometry energies may, in fact, govern the binding thermodynamics of this system. However, this conclusion cannot be made from the FEFF 3D-QSARs since these models only distinguish binding behavior on a relative basis among the inhibitors.

The regression coefficients of the van der Waal and stretch energy terms in the best FEFF 3D-QSAR models, see Tables 5 and 9, are all negative. This general form of the regression equations supports a steric fitting model for explaining enzyme–inhibitor

binding. $E_{LR,vdW}$, $E_{LR,vdW}(LL)$, and $E_L(LL)$ are increasingly negative as each of these interactions becomes increasingly stabilizing. Thus, the better (more stabilizing) each of these interactions, the more positive (better binding) is their respective contributions to ΔG. Conversely, $\Delta E_{R,stretch}$, $\Delta E_{stretch}$, and $E_{LR,stretch}$ are increasingly positive as the valence bond geometry is increasingly distorted which, in turn, produces a decrease in ΔG. The less the distortions in valence bond geometry of the ligand and receptor, the better the binding.

Electrostatics, hydrogen bonding, and solvation energetics show up in some of the four-descriptor FEFF 3D-QSAR models in Tables 5 and 9. Nevertheless, these interaction energy terms are minor descriptors relative to the van der Waal and bond stretching energy. It would seem that the hydrogen bond water bridges between the enzyme and some inhibitors, as observed by X-ray, are the only significant nonsteric binding interaction.

The GFA-MLR analysis provides a family of QSAR models for understanding the contributions of the FEFF to account for the explanation in the variance in the binding affinities of the glucose inhibitor analogue series. The dispersion energy terms, including the intramolecular vacuum energy of the unbound ligand, the intermolecular van der Waals interaction energy, and the van der Waals energy of the bound ligand account for 77% ($r^2 = 0.77$) of the variance in the biological activity over the training set. The inclusion of the FEFF solvation terms, as seen in Table 9, provide for additional explanation of the variance in inhibitory activity. Thus, for the training set of glucose analogue inhibitors of GP*b*, the solvation and dispersion energetics provide a basis for the understanding of the variance in the binding affinities.

Reality is stretched with the finding that the absolute total energy of the ligand in the unbound state, $E_L(LL)$, is a major descriptor term in the FEFF 3D-QSAR models. The explanation of this finding is uncovered by monitoring the behavior of the three related descriptors, $E_L(LL)$, the total absolute bound ligand energy, $E_{LR}(LL)$, and their difference, $\Delta E_{LR}(LL)$. During the GFA-MLR optimization all three descriptors survive for a while, but $\Delta E_{LR}(LL)$ is the first to be lost of the three in the latter phases of model optimization. A difference term of the form $aE_{LR}(LL) - bE_L(LL)$ survives in the models until very near the end of the optimization process when $E_{LR}(LL)$ is finally lost. Thus, $E_L(LL)$ is "statistically" representing the *difference* in the total bound and unbound ligand energies which does make physical sense in a ligand–receptor binding process. We stress that an investigator **not** pursue increasing inhibitory binding of new ligands by designing the ligands to be of low absolute intramolecular energy, $E_L(LL)$, through, for example, intramolecular ligand hydrogen bonding. Rather, the investigator should pursue enhancing inhibitory potency of new ligands by **minimizing the difference** between bound and unbound ligand conformational energy.

An FEFF 3D-QSAR is not a QSAR in the "classic" sense. It is a force field that can be used to predict ligand–receptor binding thermodynamics. The predictive use of this force field is in doing **virtual** screening of new hypothetical ligands. This application of the FEFF 3D-QSAR models will be one of the topics of a paper which is in preparation. However, investigators will, and probably should, inspect an FEFF 3D-QSAR for **conceptual insight** into the feature/descriptor requirements for activity. Still, investigators should not be surprised, nor disappointed, if the descriptors do not readily admit to a level of interpretation that allows specific compounds to be identified for synthesis. The dispersion energy term in the FEFF 3D-QSAR models derived in this work is an example of a descriptor that cannot be readily interpreted in terms of specific candidate compounds to synthesize but, nevertheless, is a well-defined property of a molecule and molecular systems.

The FEFF 3D-QSAR models developed in this study, when taken in comparison to FEFF binding models developed for other inhibitor–enzyme systems, suggest the FEFF methodology has discriminating capabilities to dissect the binding thermodynamic behavior of ligand–receptor systems. In the renin–peptidomimetic inhibitor study of Tokarski and Hopfinger,[5] the intrinsic stability of the ligand, $E_L(LL)$, and $E_{LR,vdW}$ are found to be major descriptors in the ΔG FEFF 3D-QSAR models. However, the change in solvation free energy upon binding as well as the changes in the binding entropies of both the ligand and the receptor are also significant descriptors in the best FEFF 3D-QSAR models.

In an unpublished FEFF 3D-QSAR analysis of a set of acetylcholinesterase analogue inhibitors to Aricept [E2020][27] (used for treatment of Alzheimer's disease), the major FEFF descriptors were found to be the ligand–receptor hydrogen-bonding energy and the ligand solvation binding energy. Overall, as observed from experimental studies on the binding thermodynamics of ligand–receptor systems,[28] different types of binding interactions are predicted by FEFF 3D-QSAR analysis to play dominant roles in different ligand–receptor systems.

Acknowledgment. This work was supported in part by DowAgro Sciences and by an NSF SBIR Phase I Grant (No. DMI-9560439) to the Chem21 Group, Inc. Resources of the Laboratory of Molecular Modeling and Design were used to perform this study. Dr. John S. Tokarski, formerly of UIC and now at Bristol Myers Squibb, gave us many helpful suggestions and insight over the course of this investigation.

References

(1) Hopfinger, A. J.; Burke, B. J. In *Concepts and Applications of Molecular Similarity*; Johnson, M. A., Maggiora, G. M., Eds.; Wiley: New York, 1990; p 173.
(2) Hopfinger, A. J.; Nakata, Y.; Max, N. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel-Dordrecht: The Netherlands, 1981; p 431.
(3) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; DeSolms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Drake, P. L.; Zuagy, J. A. A *priori* Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site. *J. Med. Chem.* **1995**, *38*, 305–317.
(4) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.
(5) Tokarski, J. S.; Hopfinger, A. J. Prediction of Ligand–Receptor Binding Thermodynamics by Free Energy Force Field (FEFF) 3D-QSAR Analysis: Applications to a Set of Peptidometic Renin Inhibitors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 792–811.

(6) Hopfinger, A. J. A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *120*, 7196–7206.
(7) Sprang, S. R.; Goldsmith, E. J.; Fletterick, R. J.; Withers, S. G.; Madsen, N. B. *Biochemistry* **1982**, *21*, 5364–5371.
(8) Hopfinger, A. J. In *Conformational Properties of Macromolecules*; Academic Press: New York, 1973; p 71.
(9) Koehler, M. G.; Hopfinger, A. J. Molecular Modeling of Polymers: 5. Inclusion of Intermolecular Energies in Estimating Glass and Crystal-Melt Transition Temperatures. *Polymer* **1989**, *30*, 116–126.
(10) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
(11) Hopfinger, A. J.; Pearlstein, R. A. Molecular Mechanics Forcefield Parameterization Procedures. *J. Comput. Chem.* **1984**, *5*, 486-492.
(12) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V_1 and V_2 Torsional terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
(13) Martin, J. L.; Johnson, L. N.; Withers, S. G. Comparison of the Binding of Glucose and Glucose 1-Phosphate Derivatives to T-State Glycogen Phosphorylase *b*. *Biochemistry* **1990**, *29*, 10745–10757.
(14) Martin, J. L.; Veluraja, K.; Ross, K.; Johnson, L. N.; Fleet, G. W. J.; Ramsden, N. G.; Bruce, I.; Orchard, M. G.; Oikonomakos, N. G.; Papageorgiou, A. C.; Leonidas, D. D.; Tsitoura, H. S. Glucose Analogue Inhibitors of Glycogen Phosphorylase: The Design of Potential Drugs of Diabetes. *Biochemistry* **1991**, *30*, 10101–10116.
(15) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Bichard, C. J. F.; Orchard, M. G.; Fleet, G. W. J.; Oikonomakos, N. G.; Leonidas, D. D.; Kontou, M.; Papageorgioui, A. Design of Inhibitors of Glycogen Phosphorylase: A Study of α- and β-C-Glucosides and 1-Thio-β-D-glucose Compounds. *Biochemistry* **1994**, *33*, 5745–5758.
(16) Watson, K. A.; Mitchell, E. P.; Johnson, L. N. Glucose Analogue Inhibitors of Glycogen Phosphorylase: from Crystallographic Analysis to Drug Prediction Using *GRID* Force-Field and *GOLPE* Variable Selection. *Acta Crystallogr.* **1995**, *D51*, 458–472.
(17) Fischer, E. H.; Krebs, E. G. *Methods Enzymol.* **1962**, *5*, 362–372.
(18) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Databank, A Computer Based Archival File for Macromolecular Structure. *J. Mol. Biol.* **1977**, *112*, 535–542.
(19) *QUANTA Version 3.3*; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1993.
(20) Doherty, D. C. *MOLSIM User Guide*; The Chem21 Group: 1780 Wilson Dr., Lake Forest, IL 60045, 1994.
(21) Pearlstein, R. A. CHEMLAB–II Users Guide, Version 11.1; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1991.
(22) Rogers, D.; Hopfinger, A. J. Applications of Genetic Function Approximation to Quantitative Structure–Activity Relationships and Quantitative Structure–Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
(23) Rogers, D. G/SPLINES: A Hybrid of Friedman's Multivariate Adaptive Regression Splines (MARS) Algorithm with Holland's Genetic Algorithm. In *The Proceedings of the Fourth International Conference on Genetic Algorithms*; San Diego, 1991; p 38.
(24) Rogers, D. Private communication, 1996.
(25) Rogers, D. Evolutionary Statistics: Using a Genetic Algorithm and Model Reduction to Isolate Alternate Statistical hypotheses of Experimental Data. In *Proceeding of the Seventh International Conference on Genetic Algorithms*; East Lansing, MI; Morgan-Kaufmann: San Francisco, CA, 1997.
(26) Rogers, D. *WOLF Reference Manual Verson 5.5*; Molecular Simulation Inc.: 1994.
(27) Kawakami, Y.; Inoue, A.; Kawai, T.; Wakita, M.; Sugimoto, H.; Hopfinger, A. J. The Rationale for E2020 As A Potent Acetylcholinesterase Inhibitor. *Bioorg. Med. Chem.* **1996**, *4*, 1429–1446.
(28) Chervenak, M. C.; Toone, E. J. A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Ligand Binding. *J. Am. Chem. Soc.* **1994**, *116*, 10533–10539.
(29) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct. Act. Relat.* **1988**, *7*, 18–25. 28

JM980515P